

文章中の重複表現の指摘方法の提案

中野 智彦[†] 丸山 広[†] 高嶋 章雄[†] 中村 太一[†]

東京工科大学[†]

1. はじめに

論理的で分かりやすい文章を書く技術は重要である。我々は、技術文章を書く能力を高めるために、日本語文法と技術文章で使われる表現に則り、分かり難い表現を指摘する校正・推敲システムを開発した[1][2][3]。本システムは学生の実験レポート作成に供されている。実験レポートを分析したところ、一文中に同じ意味や語句を繰り返す表現が多く確認された。このような重複表現は、文章が稚拙に見られるだけでなく、文意が正しく表現できない問題がある。

本稿では、文章中の重複表現を検出し指摘し、校正を支援する方法の有効性を示すことを目的とする。重複表現の原因となる、品詞の出現を抽出するルールを校正・推敲システムに実装し、学生の実験レポートに対する指摘の結果を分析し、提案方法の有効性を示す。

2. 頻出する品詞の重複表現

文章を読みにくくする要因を特定するため、学生 117 人の実験レポート（約 14000 文）を分析した結果、4 種類の重複表現が頻出することを確認した。実例を交えて詳細を以下に述べる。

【サ変接続】一文中に同じサ変接続の連続がある

サ変接続は、名詞としても動詞としても使うことの出来る品詞である。同じサ変接続の連続は、冗長な表現である[4]。

例: アンケート調査の目的の達成として、十分に達成することができたと考えられる。

【接続助詞】一文中に同じ接続助詞の連続がある

接続助詞は、前後の文節の関係を示す役割がある。そのため、同じ接続助詞の連続は、読み手に混乱をもたらすだけでなく、日本語文法として誤っている[5]。

例: キーワードが多く抽出されたが、クラス数 10 ということもあるが、逆に欠点となるようなキーワードが表示されなかった。

【代名詞】一文中に同じ代名詞の連続がある

代名詞は文脈に存在する名詞または名詞句を指す品詞である。一文中の代名詞の連続は、その代名詞が指すものを曖昧にするだけでなく、

冗長な表現になる。

例: これらのことから、SNS は、日記を他人に伝えることが出来ることが便利と考えられていると思われる。

【接続詞】一文中に同じ接続詞の連続がある

同じ接続詞の連続は、誤った日本語文法では無いが、文章を陳腐にする。

例: そして、得られた評判は、総じて旅行は楽しいという評判が、主であると考えられる。そして、家族と一緒にいくと楽しいという評判もあると考えられる。

3. 重複表現指摘手法

3.1 品詞ごとの指摘ルール

前述で述べた重複表現を自動抽出するため、各品詞ごとに指摘ルールを作成した。詳細を以下に述べる。なお、指摘する品詞の間に並列助詞を含む文は、指摘しないこととした。

表 1 指摘ルール

品詞	指摘条件	例外ルール	検査単位
サ変接続	2 回以上の使用で指摘	あり	一文
接続助詞	2 回以上の使用で指摘	あり	一文
代名詞	2 回以上の使用で指摘	なし	一文
接続助詞	2 回以上の使用で指摘	なし	前後一文

【サ変接続】

一文中に 2 回同じサ変接続が出現したとき指摘する。ただし、サ変接続の前後に名詞が含まれていた時は、意味が変わり重複表現とならないことがあるため数えない。

【接続助詞】

一文中に同じ接続助詞が出現したとき、指摘する。例外が無いと思われる接続助詞「ば」「と」「が」「のに」「ので」「から」「し」「ても（でも）」「けれど（けれども）」「て（で）」に対し指摘する。

【代名詞】

指摘する代名詞は前述の分析結果で多くの重複例が見つかった「こと」の連続使用に対し指摘する。一文中に三回以上使用で指摘する。前文に現れる名詞句を指す「こと」と、同文中のものを指す「こと」の 2 回までが、読み難くない「こと」の利用回数であるとし、3 回以上の「こと」の利用使用を指摘する。

【接続詞】

前後一文で同じ接続詞が出現したとき指摘する。

A Method for Pointing out Redundant Expressions

[†]Nakano TOMOHIKO [†]Hiroshi MARUYAMA [†]Akio TAKASHIMA [†]Taichi NAKAMURA; Graduate School of Bionics, Computer and Sciences, Tokyo University

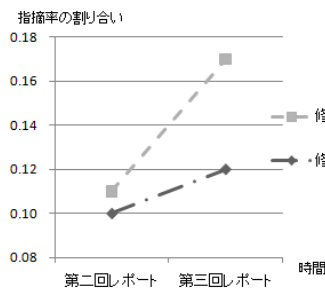


図1 全重複表現の指摘数の割合

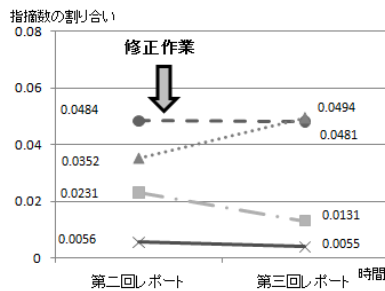


図2-1 指摘ルールごとの指摘数の割合(修正あり)

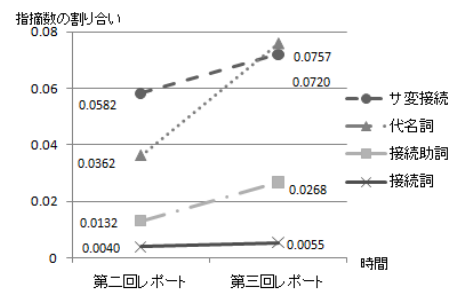


図2-2 指摘ルールごとの指摘数の割合(修正なし)

3.2 重複表現指摘の導入

著者らが作成した校正システム[3]に前述の指摘ルールを導入した。重複表現の指摘の流れを図2に示す。

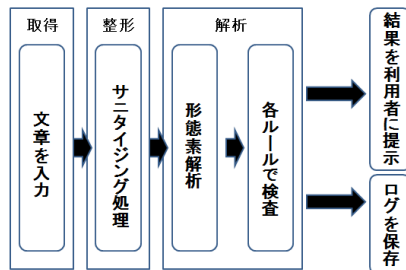


図2 重複表現の指摘の流れ

4. 評価実験

4.1 実験概要

指摘ルールを導入した校正システムの利用後の学習効果を評価するために、東京工科大学の実験講義(2010年度後期)にシステムを導入した。使用したデータは、11月の第二回レポートと1カ月後の第三回レポートの指摘数の増減を分析した。学生により分量に差があるため、指摘数を正規化し、一文あたりの指摘数の割合を分析に利用した。

4.2 評価結果

45名の学生がシステムを568回利用した。述べ102,706文が対象となり、8,069個の重複表現が指摘された。

第二回のレポートにおいて、重複表現の修正を行った人は24名、修正を行わなかった人は21名だった。第二回のレポートを修正した人、修正しなかった人の第二回レポートと第三回レポートの全重複表現の指摘数の割合の平均を図1に示す。また、4種類の指摘ごとの指摘数の割合をそれぞれ、図2-1、図2-2に示す。

5. 考察

図1から、修正の有無に関わらず指摘数の割合が増加しており、レポートの性質により重複

表現の出現率が変わると考えられる。また、修正ありの人は、修正なしの人に比べ指摘数の割合の増加値が小さい。このことから、校正システムの利用が、指摘数の割合に影響を与えたと言える。

図2-2より、修正無しの人は指摘数の割合がすべて増加している。一方、図2-1では、修正ありの人は、代名詞の指摘を除き指摘数の割合が減少している。代名詞の指摘数の割合は、修正なしの人が209%増加、修正ありの人は143%である。重複表現が出やすい第三回レポートにおいて、指摘数の割合の増加値が抑えられた。よって、代名詞の指摘の効果があったと確認できる。以上のことから、本校正システムの利用者は、指摘事項の学習により、指摘数が減少したと結論付ける。

6. おわりに

本研究では、文の分かりやすさを向上させる手法として、重複表現を指摘する方法を提案した。学生のレポート作成の際にシステムを提供し、学習効果を確認した。現状の指摘アルゴリズムでは、重複表現では無いにも関わらず、重複表現を検出してしまうケースがあった。サ変接続は名詞表現と動詞表現が混在しているときのみ指摘すべきである。また、現在のアルゴリズムは、代名詞の指摘は指示代名詞のみを対象としているが、人称代名詞も重複表現として指摘できると考えている。

参考文献

- [1] 大長達也, 丸山広, 中村太一: “係り受け構造を用いた読みやすい文章への推敲支援”, 情報処理学会第69回全国大会, 6Q-6 pp.2-433 - 2-434, 2007
- [2] 須藤崇志, 丸山広, 中村太一: “文を分かりにくくする要因の分析と改善支援手法の提案”, 電子情報通信学会, 2008
- [3] 坂本俊介, “文章の誤りを指摘する文章校正支援手法の提案”
- [4] 永山 壽昭: 文章・表現 200の鉄則 日経 BP 社
- [5] 一ノ坪 俊一: 書く技術 日本経済新聞社